

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 21/02 (2006.01)



[12] 发明专利说明书

专利号 ZL 03818622.5

[45] 授权公告日 2007 年 11 月 14 日

[11] 授权公告号 CN 100349209C

[22] 申请日 2003.3.28 [21] 申请号 03818622.5

[30] 优先权

[32] 2002. 6. 3 [33] US [31] 10/158,908

[86] 国际申请 PCT/US2003/009538 2003. 3. 28

[87] 国际公布 WO2003/102924 英 2003. 12. 11

[85] 进入国家阶段日期 2005. 2. 2

[73] 专利权人 英特尔公司

地址 美国加利福尼亚州

[72] 发明人 亚历克斯·洛佩斯-埃斯特拉达

[56] 参考文献

US 5825320 A 1998. 10. 20

CN 1031607C 1996. 4. 17

US 6345125 B2 2002. 2. 5

SCALABLE EMBEDDED ZERO TREE
WAVELET PACKET AUDIO CODING. PAO.
CHI, CHANG ET AL, ET AL, AL. 2001 IEEE
THIRD WORKSHOP ON SIGNAL PROCESSING
ADVANCES IN WIRELESS COMMUNICATIONS.
2001

审查员 刘红梅

[74] 专利代理机构 北京嘉和天工知识产权代理事
务所

代理人 严 慎

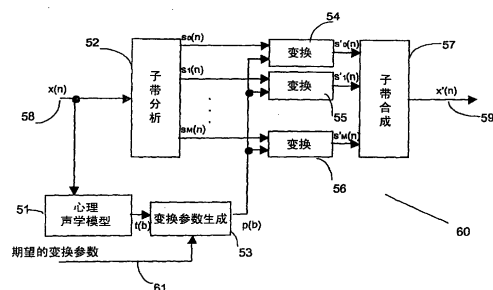
权利要求书 2 页 说明书 6 页 附图 5 页

[54] 发明名称

数字音频信号的感知标准化方法及标准化器

[57] 摘要

将接收到的数字音频数据标准化的方法，包括将数字音频数据分解到多个子带中，并将心理声学模型应用于所述数字音频数据，以生成多个掩蔽阈值。该方法还包括基于掩蔽阈值和期望的变换参数生成多个变换调整参数，并将所述变换调整参数应用于多个子带，以生成变换后的子带。



1. 一种将接收到的数字音频数据标准化的方法，包括：
将所述数字音频数据分解到多个子带中，
将心理声学模型应用于所述数字音频数据，以生成多个掩蔽阈值；
基于所述掩蔽阈值和期望的变换参数，生成多个变换调整参数；以及
将所述变换调整参数应用于所述子带，以生成变换后的子带。
2. 如权利要求1所述的方法，其中，所述多个子带中的每一个都对应于所述心理声学模型的多个临界频带中的一个临界频带，并且其中所述掩蔽阈值是所述多个临界频带的函数。
3. 如权利要求1所述的方法，还包括：
合成所述变换后的子带，以生成标准化数字音频数据。
4. 如权利要求1所述的方法，其中所述接收到的数字音频数据包括多个数字块。
5. 如权利要求1所述的方法，其中所述数字音频数据是基于小波包树而被分解的。
6. 如权利要求1所述的方法，其中所述心理声学模型包括听觉的绝对阈值。
7. 如权利要求2所述的方法，其中所述多个变换调整参数是通过提供子带支配性度量而生成的。
8. 一种标准化器，包括：
子带分析模块，所述子带分析模块将接收到的数字音频数据分解到多个子带中；
心理声学模型模块，所述心理声学模型模块将心理声学模型应用于所述接收到的数字音频数据以生成多个掩蔽阈值；
变换参数生成模块，所述变换参数生成模块基于所述掩蔽阈值和期望变换参数来生成多个变换调整参数；以及
多个子带变换模块，所述多个子带变换模块将所述变换调整参数应用于所述子带，以生成变换后子带。
9. 如权利要求8所述的标准化器，其中，所述多个子带中的每一个都对应于所述心理声学模型的多个临界频带中的一个临界频带，并且其中所述掩蔽阈值是所述多个临界频带的函数。
10. 如权利要求8所述的标准化器，还包括：
子带合成模块，所述子带合成模块合成所述变换后的子带，以生成标准化数字音频数

据。

11. 如权利要求 8 所述的标准化器,其中所述接收到的数字音频数据包括多个数字块。
12. 如权利要求 8 所述的标准化器,其中所述数字音频数据是基于小波包树而被分解的。
13. 如权利要求 8 所述的标准化器,其中所述心理声学模型包括听觉的绝对阈值。
14. 如权利要求 9 所述的标准化器,其中所述多个变换调整参数是通过提供子带支配性度量而生成的。

数字音频信号的感知标准化方法及标准化器

技术领域

本发明的一种实施方案涉及数字音频信号。更具体地说，本发明的一种实施方案涉及数字音频信号的感知标准化（perceptual normalization）。

背景技术

数字音频信号常常被标准化以解决条件的变化或用户偏好的变化。标准化数字音频信号的例子包括改变信号的音量或改变信号的动态范围。何时需要改变动态范围的一个例子是当 24 位编码数字信号必须被转换成 16 位编码数字信号以适应 16 位放音设备。

数字信号的标准化经常被盲目地执行于数字音频源，而不考虑它的内容。多数情况下，由于信号的所有分量都被同样地改变，所以盲目的音频调整将产生感知上可察觉的人工成份（artifact）。数字音频标准化的一种方法包括对输入的音频信号施加函数变换来压缩或扩展数字信号的动态范围。这些变换本质上可以是线性的或非线性的。然而，最常用的方法却是利用输入音频的点对点的线性变换。

图 1 是说明一个示例的图，在该示例中，对正态分布的数字音频样本施加线性变换。这种方法不考虑隐藏在信号中的噪声。通过应用增大信号平均值和范围的函数，隐藏在信号中的附加噪声也将被放大。例如，如果图 1 所示的分布对应于某种误差或噪声分布，那么应用简单的线性变换将导致更高的平均误差和伴行的更宽的范围，如通过比较曲线 12（输入信号）和曲线 11（标准化信号）所示的那样。这在大多数音频应用中是典型的不良状况。

根据上述内容，需要一种改进的用于数字音频信号的标准化技术，它可以降低或消除感知上可察觉的人工成份。

附图说明

图 1 是说明一个示例的图，在该示例中，对正态分布的数字音频样本施加线性变换。

图 2 是说明掩蔽信号频谱的假想示例的图。

图 3 是根据本发明一种实施方案的标准化器的功能块的框图。

图 4 是说明小波包树形结构（Wavelet Packet Tree Structure）的一种实施方案的图。

图 5 是可用于实现本发明的一种实施方案的计算机系统的框图。

具体实施方式

本发明的一种实施方案是一种通过基于听觉系统的特性分析数字音频数据，来有选择地改变音频分量的性质，从而标准化该数字音频数据的方法。在一种实施方案中，该方法包括将音频数据分解到多个子带中，以及将心理声学模型应用于该数据。结果，防止了感知上可察觉的人工成份的引入。

本发明的一种实施方案利用感知模型和“临界频带”(critical bands)。听觉系统经常被建模为滤波器组(filter bank)，该滤波器组将音频信号分解到称为临界频带的多个频带中。临界频带由一个或多个音频分量组成，该音频分量被视为单个实体。某些音频分量能够掩蔽一个临界频带内的其他分量(内掩蔽)以及来自其他临界频带的分量(互掩蔽)。尽管人类的听觉系统高度复杂，但是计算模型已经成功地用于许多应用。

感知模型或心理声学模型(Psycho-Acoustic Model, “PAM”)通常按照声压级(Sound Pressure Level, “SPL”)来计算阈值掩蔽(threshold mask)，所述阈值掩蔽作为临界频带的函数。下降到阈值边缘以下的任何音频分量都将被“掩蔽”，因而听不见。糟糕的位速率降低或音频编码算法利用这个现象将量化误差隐藏到这个阈值以下。因此，应该注意设法不暴露这些误差。如上结合图1所描述的那样，简单的线性变换将潜在地放大这些误差，使用户可以听见它们。另外，来自A/D转换的量化噪声可能会通过动态范围扩展过程而暴露出来。另一方面，如果发生简单的动态范围压缩，则高于阈值的可听信号会被掩蔽。

图2是说明掩蔽信号频谱的假想示例的图。阴影区域20和21对于普通听众是听得见的。下降到掩蔽22之下的任何东西都是听不见的。

图3是根据本发明一种实施方案的标准化器60的功能块框图。图3中模块的功能性可以由硬件组件，由处理器执行的软件指令，或者由硬件或软件的任意组合来完成。

在输入58处接收进入的数字音频信号。在一种实施方案中，数字音频信号的形式是长度为N的输入音频块， $x(n)$ $n=0, 1, \dots, N-1$ 。在另一种实施方案中，数字音频信号的整个文件可由标准化器60来处理。

子带分析模块52从输入58接收到数字音频信号。在一种实施方案中，子带分析模块52将长度为N的输入音频块 $x(n)$ $n=0, 1, \dots, N-1$ 分解到M个子带中， $s_b(n)$ $b=0, 1, \dots, M-1$ ， $n=0, 1, \dots, N/M-1$ ，其中每个子带都与一个临界频带相关联。在另一种实施方案中，子带与任何临界频带都不相关联。

在一种实施方案中，子带分析模块52使用基于小波包树(Wavelet Packet Tree)的子带分析方案。图4是说明小波包树形结构的一种特定实施方案的图，假定在44.1KHz对输入音频进行采样，该树形结构由29个输出子带组成。图4所示的树形结构随采样率而变

化。每条线都表示一分为二的抽选（低通滤波器，后面跟着因子为 2 的亚采样（sub-sampling））。

在子带分析期间使用的低通小波滤波器的实施方案可以随优化参数而变，这取决于感知音频质量和计算性能之间的折衷。一种实施方案利用 N=2 的 Daubechies 滤波器（通常所说的 db2 滤波器），该滤波器的标准化系数由序列 c[n] 给出：

$$c[n] = \left\{ \frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}} \right\}$$

每个子带都试图与人类听觉系统的临界频带具有相同的中心。因此，可以使心理声学模型模块 51 和子带分析模块 52 的输出之间具有简单直接的关联。

心理声学模型模块 51 也从输入 58 接收数字音频信号。心理声学模型（“PAM”）利用算法建立人类听觉系统的模型。已知很多不同的 PAM 算法可以用于本发明的实施方案。然而，对于大多数算法而言，其理论基础是一样的：

- 将音频信号分解到频谱域中—快速傅立叶变换（“FFT”）是最广泛应用的工具。
- 将频谱带分组成临界频带。这是从 FFT 采样到 M 个临界频带的映射。
- 确定临界频带内的音调和非音调（像噪声的分量）。
- 通过使用能级、音调和频率位置，计算每个临界频带分量的单独掩蔽阈值。
- 计算作为临界频带的函数的某类掩蔽阈值。

PAM 模块 51 的一种实施方案利用听觉的绝对阈值（或静音阈值），以避免与更复杂的模型相关的高计算复杂度。根据声压级（或功率谱的对数），通过下面的方程给出听觉的最小阈值：

$$T(SPL) = 3.64f^{-0.8} - 6.5e^{[-0.6(f-3.3)^2]} + 0.001f^4 \quad (1)$$

其中 f 以千赫给出。

从千赫频率到临界频带（或巴克率，bark rate）的映射通过下面的方程来实现：

$$f_b = 13 \arctan(0.76f) + 3.5 \arctan(f/7.5)^2 \quad (2)$$

$$BW(Hz) = 15 + 75[1 + 1.4f^2] \quad (3)$$

其中 BW 是临界频带的带宽。由频率线 0 开始并且建立临界频带，使得一个频带的上沿是下一个频带的下沿，可以累积方程（1）中听觉的绝对阈值的值，使得：

$$T(b) = \frac{1}{N_b} \sum_{\omega=\omega_l}^{\omega_h} 10^{\frac{T(SPL)}{10}} \quad (4)$$

其中 N_b 是临界频带中频率线的数量， ω_l 和 ω_h 是临界频带 b 的下沿和上沿。

在这种实施方案中，在 N 个输入样本的重叠块上计算输入音频的实数值 FFT（real valued FFT）；由于实数值信号的 FFT 的对称性，N/2 条频率线被保留下来。然后，如下

计算输入音频的功率谱：

$$P(\omega) = \text{Re}(\omega)^2 + \text{Im}(\omega)^2 \quad (5)$$

接着，将信号的功率谱和掩蔽阈值（在这种情况下是静音阈值）传递到下一个模块。将 PAM 模块 51 的输出被输入到变换参数生成模块 53 中。变换参数生成模块 53 在输入 61 处接收基于期望的标准化或变换的期望变换参数作为输入。在一种实施方案中，根据掩蔽阈值和期望的变换，变换参数生成模块 53 生成作为临界频带函数的动态范围调整参数， $p(b)$ $b=0, 1, \dots, M-1$ 。

在一种实施方案中，变换参数生成模块 53 首先试图根据更具支配性的临界频带在音量和掩蔽方面的特性来提供这些频带的定量测定。这种定量测定被称为“子带支配性度量”（Sub-band Dominancy Metric, “SDM”）。因此，为了在可能会隐藏噪声或量化误差的非支配性频带的变换中减少侵略性，对所述动态范围标准化参数进行“按摩（massage）”处理。

SDM 被计算为特定临界频带内的频率线与相关掩蔽阈值之间的绝对差的和：

$$SDM(b) = \text{MAX}[P(\omega) - T(b)] \quad \omega = \omega_l \rightarrow \omega_h \quad (6)$$

其中 ω_l 和 ω_h 对应于临界频带 b 的下频率限和上频率限。

因此，那些 $P(\omega)$ 明显大于掩蔽阈值的临界频带被认为是支配性的（dominant），这些临界频带的 SDM 趋近于无穷大，而 $P(\omega)$ 下降到掩蔽阈值之下的临界频带则是非支配性的，这些临界频带的 SDM 趋近于负无穷大。

为了将 SDM 度量限制在 0.0 到 1.0 的范围内，可以利用下面的方程：

$$SDM'(b) = \frac{1}{\pi} a \tan(SDM(b) / \gamma - \delta) + \frac{1}{2} \quad (7)$$

其中参数 γ 和 δ 根据应用来优化，如： $\gamma=32$ ， $\delta=2$ 。

变换参数生成模块 53 除了生成 SDM 度量之外，还修改期望的输入变换参数 61。在一种实施方案中，假定将对输入信号数据执行

$$x'(n) = \alpha x(n) + \beta \quad (8)$$

形式的线性变换。参数 α 和 β 可以由用户/应用提供的，或者是根据音频信号统计量自动计算出来的。

作为变换参数生成模块 53 操作的实施例，假定期望对值在 -32768 到 32767 范围内的 16 位音频信号的动态范围进行标准化。在一种实施方案中，所有处理过的音频都将被标准化到由 $[ref_min, ref_max]$ 指定的范围。在一种实施方案中， $ref_min=-20000$ ， $ref_max=20000$ 。导出变换参数的自动方法可以是：

·计算初始样本块中的最大和最小信号值。

·确定参数 α 和 β ，使得变换块的新的最大和最小值被标准化到 $[-20000, 20000]$ 。这可以利用初等代数通过确定线的斜率和截距而求出：

$$\alpha = \frac{[ref_max - ref_min]}{max - min} = \frac{[20000 - (-20000)]}{max - min}$$

$$\beta = ref_max - \alpha \cdot max = 20000 - \alpha \cdot max \quad (9)$$

·对于每个进入的块以迭代的方式进行重复，同时保存在先块的 max 和 min 历史数据。

一旦确定了标准化参数，就根据 SDM 调整这些参数。对于每个子带：

$$\alpha'(b) = (\alpha - 1) \cdot SDM'(b) + 1$$

$$\beta'(b) = \beta \cdot SDM'(b) \quad (10)$$

因此，如果用于特定子带的 SDM 等于 0，如非支配性的子带，则斜率等于 1.0，并且截距等于 0。这导致不变的子带。如果 SDM 等于 1.0，如支配性的子带，则斜率和截距将等于从方程 (9) 获得的原始值。对于本实施方案来说，将被传递到标准化器 60 的子带变换模块 54-56 的参数 $p(b)$ 是 $\alpha'(b)$ 和 $\beta'(b)$ 。

子带分析模块 52 和变换参数生成模块 53 的输出被输入给子带变换模块 54-56。子带变换模块 54-56 将从变换参数生成模块 53 接收到的变换参数应用于从子带分析模块 52 接收到的各个子带。该子带变换（在如方程(8)所示的线性变换的实施方案中）由下面的方程表示：

$$s'_b(n) = \alpha'(b)s_b(n) + \beta'(b) \quad b = 0, 1, \dots, M-1; n = 0, 1, \dots, N/M-1 \quad (11)$$

在一种实施方案中，子带变换模块 54-56 的输出是标准化器 60 的最终输出。在本实施方案中，数据可以随后馈入编码器，或者可以被分析。

在另一种实施方案中，子带变换模块 54-56 的输出由子带合成模块 57 接收，该模块 57 将合成变换后的子带 $s'_b(n)$ $b=0, 1, \dots, M-1, n=0, 1, \dots, N/M-1$ ，以在输出 59 处形成输出标准化信号 $x'(n)$ 。在一种实施方案中，由子带合成模块 57 进行的子带合成可以通过倒序图 4 所示的小波树形结构，并且改成使用合成滤波器来完成。在一种实施方案中，该合成滤波器是 $N=2$ 的 Daubechies 小波滤波器（通常所说的 db2），该滤波器的标准化系数由下面的序列 $d[n]$ 给出：

$$d[n] = \left\{ \frac{1-\sqrt{3}}{4\sqrt{2}}, \frac{-3+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{-1-\sqrt{3}}{4\sqrt{2}} \right\}$$

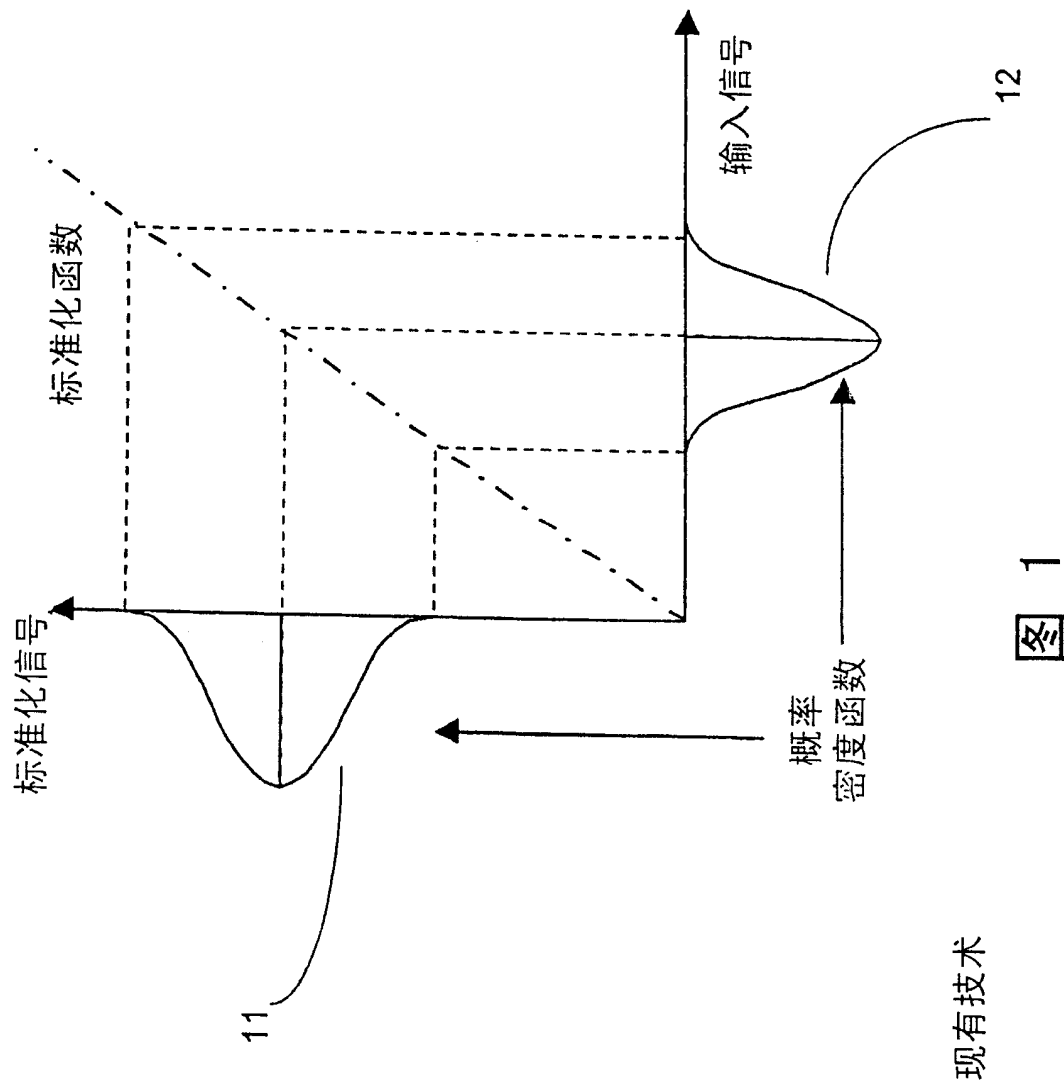
因此，利用互补小波滤波器以内插操作（上采样和高通滤波器）来代替每个抽选操作。

图 5 是可用于实施本发明一种实施方案的计算机系统 100 的框图。计算机系统 100 包括处理器 101、输入/输出模块 102 和存储器 104。在一种实施方案中，上述功能作为软件

存储在存储器 104 中，并由处理器 101 执行。在一种实施方案中，输入/输出模块 102 接收图 3 的输入 58 并输出图 3 的输出 59。处理器 101 可以是任何类型的通用或专用处理器。存储器 104 可以是任何类型的计算机可读介质。

如上所述，本发明的一种实施方案是这样一种标准化器，它实现了数字音频信号的时域变换，同时防止引入可听到的明显人工成份。各实施方案是利用人类听觉系统的感知模型来完成变换。

这里具体说明和/或描述了本发明的几种实施方案。然而可以理解，在不脱离本发明的实质和预定范围的情况下，本发明的多种修改和变化被上面的教导所覆盖，并且落在所附权利要求书的范围内。



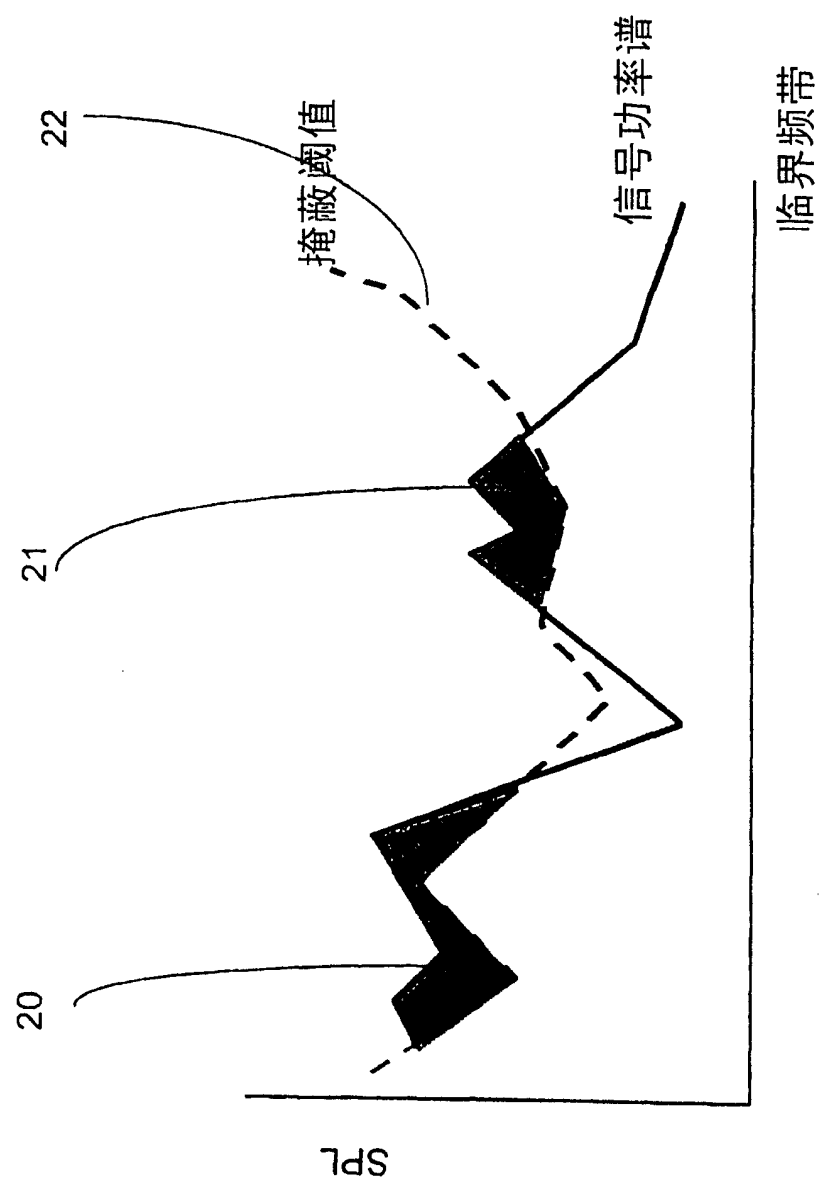


图 2

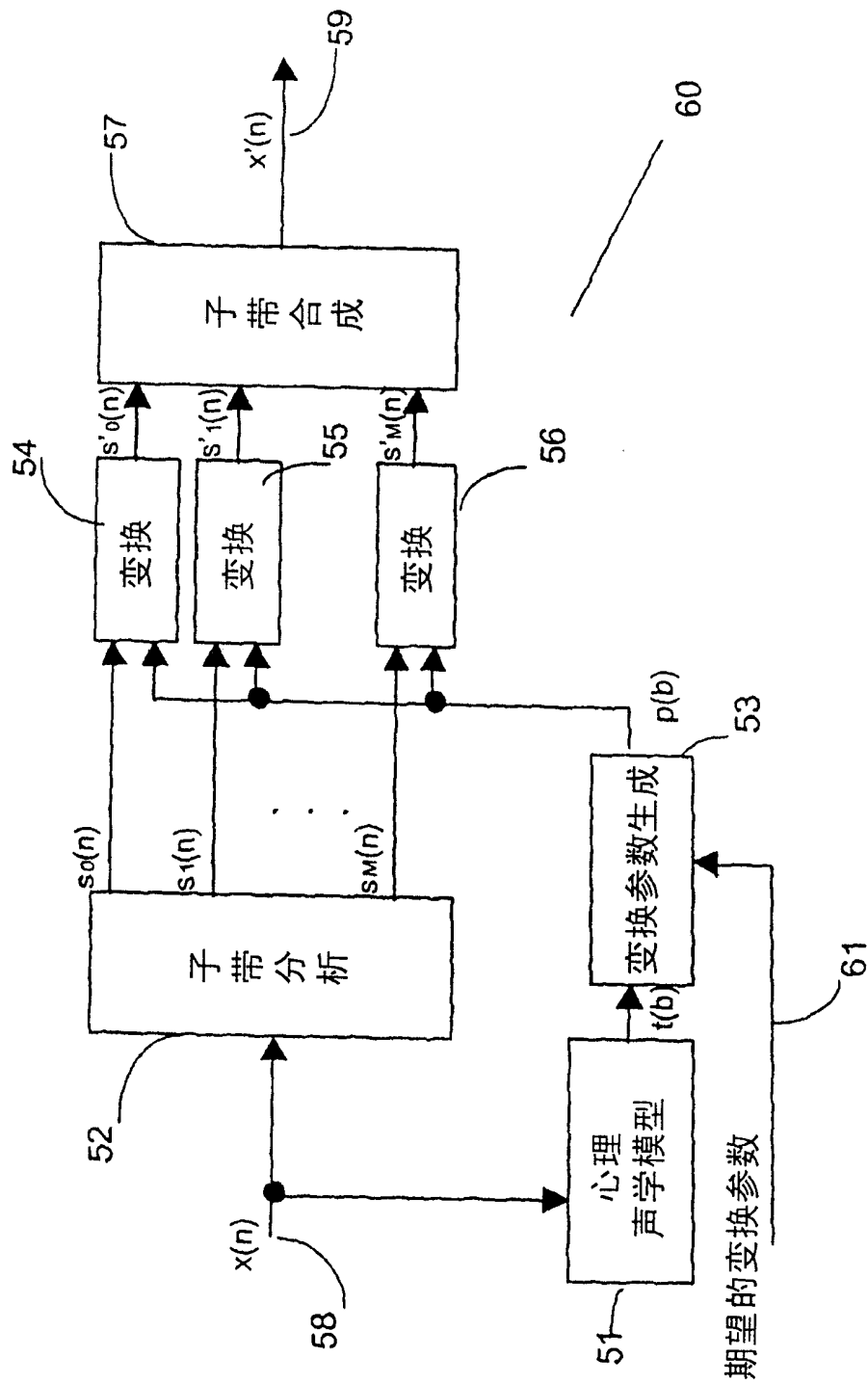


图 3

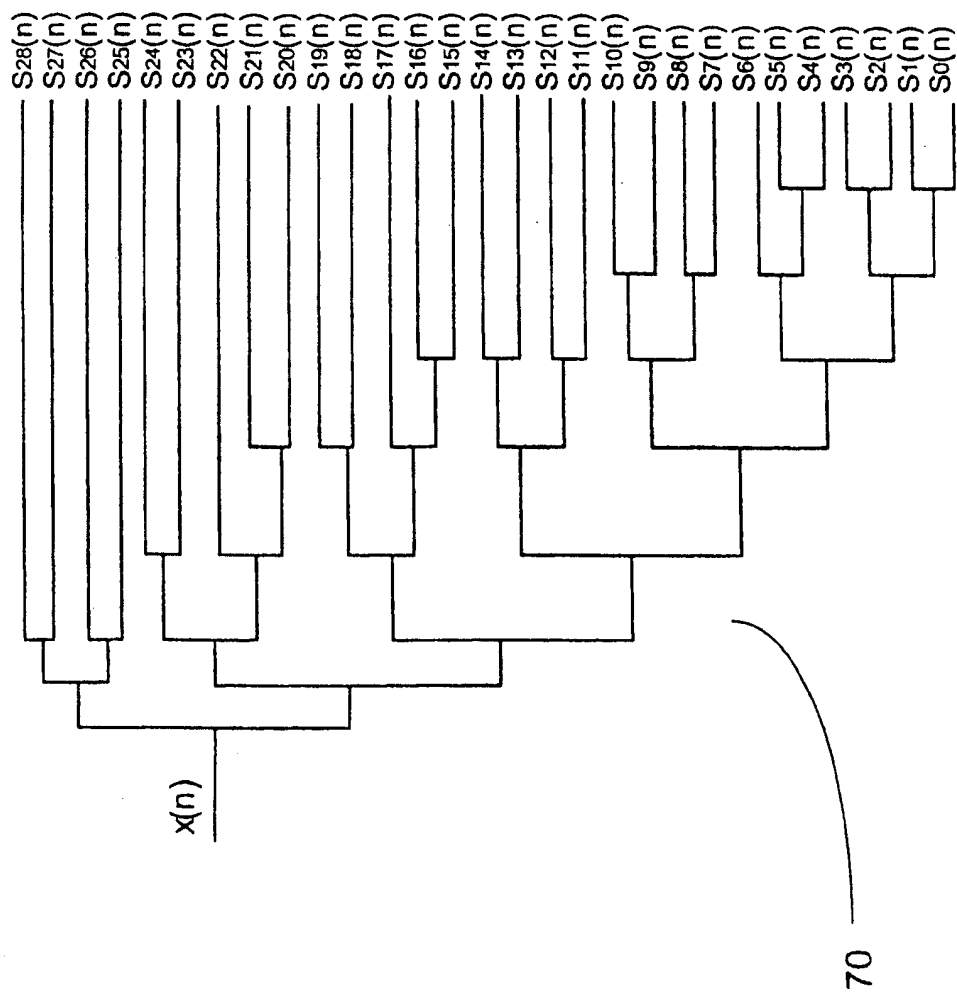


图 4

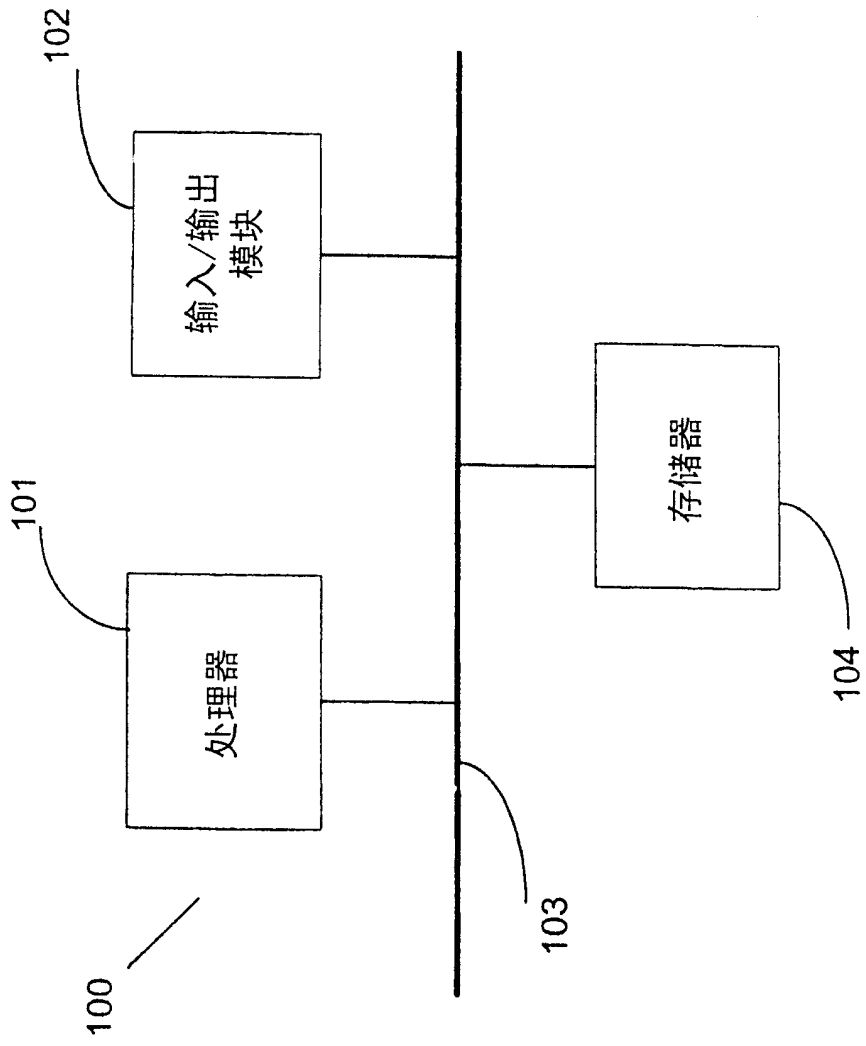


图 5